4.3 Comparison of Surveillance Data with EMFAC2000 Predictions for Emissions Regime

The following section details an analysis performed by Sierra Research (under contract to CARB) to determine how the distribution of vehicles by emission regime from the CALIMFAC model compared to data from newer surveillance data. The purpose of this analysis was to identify vehicle technology groups/pollutant combinations that should be revised.

The distribution of vehicles among emissions regimes found in the surveillance data was compared to the distributions predicted by CALIMFAC model. This was done using a chi-squared ($\chi^{2)}$ test for different population distributions. The data used were from CARB surveillance projects run between 1987 and 1994.[1] Only passenger car data were used in the comparison.

The χ^2 test was done for the following six technology groups, which had the largest sample sizes in the ARB surveillance data:

- Old technology group 6 - 1977+ Oxidation catalyst with secondary air
- Old technology group 8 - 1981+ TBI/Carb, single-bed TWC, 0.7 NOx
- Old technology group 9 - 1981+ TBI/Carb, double-bed TWC, 0.7 NOx
- Old technology group 11 - 1981+ MPFI, TWC, 0.7 NOx
- Old technology group 13 - 1981+ MPFI, TWC, 0.4 NOx

4.3.1 Test Procedures

The following test statistic was used for the χ^2 test:[3]

$$u = \sum_{i=norm}^{super} \frac{(n_i - n_{e,i})^2}{(n_{e,i})} \cdots\cdots\cdots [4-1]$$

Where the summation index, i, is taken over the five emissions regimes (normal, moderate, high, very high and super); n_i is the number of surveillance vehicles found to be in a given regime; and $n_{e,i}$ is the number expected in the regime. The expected number was found by multiplying the total number in the sample ($n = \sum_i n_i$) by the CALIMFAC predicted fraction of vehicles in the regime. In general $n_{e,i}$ is a non-integer value. This statistic and the use of the χ^2 distribution presumed a "large" sample size, but the χ^2 approximation is surprisingly good for small n when the number of regimes is greater than two (as it was here).

---

[1] These surveillance tests have the following project names in the ARB data base: 2S87C1, 2S88C1, 2S89C1, 2S89C2, 2S91C1, 2S91C2 and 2S93C1.

The statistical test to determine if the observed distribution is different from the expected distribution is based on the null hypothesis that the two distributions are the same. This hypothesis is rejected if the computed test statistic is greater than a critical value determined by the desired significance level and the degrees of freedom. For this test, the significance level represented the probability that the null hypothesis would be rejected if the observed and expected distributions were, in fact, the same. The number of degrees of freedom equals the number of regimes minus one, which was four in this calculation. For the level of significance chosen, the critical value of the test statistic, u_{crit}, is found from tables of the χ^2 distribution. The statistical test is defined as follows:

$u > u_{crit}$ Reject null hypothesis; distributions assumed different.

$u \leq u_{crit}$ Accept null hypothesis; distributions assumed the same.

The test was done using a significance level of 0.05; the critical value of the test statistic is 9.488 for this significance level.

For small sample sizes, an alternative test can be used. This test is based on the likelihood ratio, λ, defined in terms of the actual and expected number of vehicles by the following equation:

$$\lambda = \prod_{i=norm}^{\sup er} \left(\frac{n_{e,i}}{n_i} \right)^{n_i} \ldots \qquad [4-2]$$

n_i and $n_{e,i}$ have been defined previously; Π is the continued product operator. The small-sample approach is outlined below.

1. For a given sample size, n, and EMFAC2000 expected values, compute the value of λ and the probability for each possible distribution of the (integer) $\{n_i\}$.

2. Tabulate the values of λ for various distributions and the associated probability of that distribution in order of increasing λ.

3. Compute the cumulative probability for each value of λ in the table (i.e., the probability that λ will be less than or equal to a given value).

4. Select the critical value of λ as the one whose cumulative probability has the desired significance level (0.05 in these calculations).

5. Compare the value of λ for the observed distribution in the surveillance data with the critical value of λ determined in step 4. If the observed value is less than the critical value, reject the null hypothesis that the distributions are the same.

This small sample procedure was used for sample sizes of one to twelve. The large-sample statistic, u, was also computed for these samples. In general, the small-sample procedure showed lower probabilities that the observed sample was different from the EMFAC2000 distributions than the large-sample procedure. Thus, any possible error in switching from the small-sample procedure to the large-sample procedure for samples of eight or more would reject the null hypothesis when it might otherwise be accepted by the more accurate procedure.

To compare observed and expected data, the surveillance data in which vehicle age is not recorded, but odometer readings are, were compared to the EMFAC2000 predictions of emission regimes, which are calculated based on odometer mileage but are reported as a function of vehicle age. These equivalencies are shown in Table 4-10.

For a given test year,CALIMFAC provides results only for technology groups whose age is less than the maximum possible age for the technology group (e.g., predictions for a 1987 test date for technology groups that start in 1981 will be available only for vehicle ages of 1 to 6 years). There were some surveillance data that could not be used because they had accumulated high mileage in a short time. Other vehicles did not have CALIMFAC distributions because they were too old to be included in the sales fraction. This was especially true of the oxidation catalyst vehicles in old technology group 6.

Table 4-10 <u>Equivalencies Between Odometer Reading in Surveillance Data and Vehicle Age in EMFAC2000 Predictions</u>

Surveillance Data Odometer Readings (miles)	EMFAC2000 Vehicle Age Range (years)
0 to 22,000	1 and 2
22,000 to 45,000	3 and 4
45,000 to 65,999	5 and 6
66,000 to 85,000	7 and 8
Greater than 85,000	9 and greater

The initial tests were performed by subdividing the surveillance data by test year and vehicle odometer reading for each technology group. There were eight test years and five odometer/age classifications, giving 40 possible distributions of vehicles among emission regimes for each pollutant in each technology group. Because some test-year/age classifications had no vehicle data and other high-mileage vehicles in early test years had no CALIMFAC projections, there were only 101 different distributions for a given pollutant: 6 for old technology group 6, 8 for old group 13, 21 for old group 9, and 33 for old groups 8 and 11. Considering all pollutants, there were 303 (3 x 101) distributions from surveillance data to be compared to CALIMFAC data. This was the most disaggregated set of data considered.

4.3.2 Results

The χ^2 test (or the small-sample alternative) was applied to each of the 303 distributions for a specified pollutant, test year, technology group and odometer/age group. Of the 303 possible tests, 260 satisfied the null hypothesis at the 0.05 level. The distributions that failed the χ^2 test are summarized in Table 4-11. In addition to showing the test statistic for the failed distributions, Table 2-2 shows the probability that the observed test statistic would be observed if the underlying distributions were, in fact, the same. For the small-sample procedure, only this probability is shown.

Table 4-11
Description of One-year Data Sets Not Satisfying Null Hypothesis of χ^2 Test at 0.05 Significance Level
Not Shown: 260 of 303 Data Sets Satisfying Null Hypothesis

Species	Old Technology Group	Test Date	Mileage Group	Sample Size	χ^2 Result	
					Test Statistic	Proba-bility
HC	6	1993	85k+	2	-	.0067
	8	1989	85k+	62	36.53	$4x10^{-7}$
		1990	85k+	7	-	.0050
		1993	66-85k	8	-	.0093
		1993	85k+	16	16.69	.0014
		1994	66-85k	3	-	.0049
	9	1987	45-66k	9	-	.0465
		1989	85k+	10	-	.0002
		1990	22-45k	13	9.83	.0434
		1990	45-66k	6	32.37	$2x10^{-6}$
		1991	85k+	3	-	.0098
		1992	85k+	6	-	.0088
	11	1989	66-85k	19	10.75	.0295
		1989	85k+	34	13.72	.0082
		1992	66-85k	29	11.68	.0199
		1993	45-66k	4	-	.0095
		1993	85k+	16	12.98	.0114
	13	1992	22-45k	8	-	.0233
CO	6	1988	85k+	2	-	.0067
		1993	85k+	2	-	.0026

	8	1989 1990 1993 1994	85k+ 85k+ 85k+ 66-85k	62 7 16 3	18.33 - 15.97 -	.0011 .0023 .0031 .0476
	9	1988 1989 1990 1991 1992	22-45k 85k+ 45-66k 85k+ 85k+	43 10 13 3 6	10.37 - 27.09 - -	.0346 $6x10^{-6}$ $2x10^{-5}$ .0108 .0342
	11	All pass for this technology group and pollutant.				
	13	All pass for this technology group and pollutant.				
NOx	6	1989	85k+	12	-	.0042
	8	1993	45-66k	6	-	.0074
	9	1987 1989 1990 1990 1992	0-22k 85k+ 22-45k 85k+ 85k+	4 10 5 6 6	- - - - -	.0152 .0220 .0071 .0476 .0360
	11	1989 1992 1992 1992 1993 1994 1994	85k+ 45-66k 66-85k 85k+ 66-85k 66-85k 85k	34 24 29 28 5 5 5	33.30 16.49 19.85 12.81 - - -	$1x10^{-6}$ .0024 .0005 .0122 .0113 .0029 .0022
	13	All pass for this technology group and pollutant.				
Note: The "probability" column entry represents the probability that the test statistic would be observed if the two distributions were actually the same. The test statistic column contains the result of the computation in equation [1] where the total sample was greater than 12 vehicles.						

The largest number of mismatches are in old group 9, for all pollutants, old technology group 8 for HC and CO, and old technology group 11 for HC and NOx.

Many of the distributions that did not satisfy the null hypothesis at the 0.05 significance level had small sample sizes, but one of the failed distributions had a sample size of 62, one of the larger sample sizes for this test series. Twenty-four of the 43 failed distributions were for the 85,000-mile-plus category. This category in the surveillance data fleet was compared with the nine-year-plus vehicle age in the EMFAC2000 data.

One possible reason that a large number of the distributions not satisfying the null hypothesis are in 85,000-and-above mileage range is that the actual distribution of vehicles in the surveillance fleet differs from the standard distribution predicted from EMFAC2000. To illustrate this difference, the proportion of super emitters for CO emissions from old technology group 8 in CALIMFAC was examined for each of the age ranges used to classify the odometer data in the surveillance fleet. The results are shown in Table 4-12.

Table 4-12
Proportion of Super Emitters in CALIMFAC
for CO in Old Technology Group 8

Age Range (years)	1-2	3-4	5-6	7-8	9+
Range for Supers	0	0 to .007	.007 to .0205	.0205 to .0326	.0326 to .0975

The range for the proportion of super emitters is relatively broad in the nine-year-plus age category compared to the other age bins. Thus, it is possible that this age classification does not provide a good match to the odometer data range for this regime.

The data for each technology group were further aggregated into four two-year groups based on test dates (1987-1988, 1989-1990, 1991-1992, and 1993-1994). Each of these groups should represent vehicles that have undergone a different number of visits to an I&M station in the biennial program. Each technology group could provide a maximum of 15 distributions (3 two-year blocks x 5 five-mileage bins). Because of some groups with no data, there were only 5 distributions for old technology group 6, 17 for old group 8, 12 for old group 9, 19 for old group 11, and 5 for old group 13. This gave a total of 58 distributions for each pollutant, or a total of 174 distributions for comparison to CALIMFAC data. At the 0.05 level, 142 of these satisfied the null hypothesis that the distributions are similar. Table 4-13, with the same format as Table 4-11, shows the information on the distributions that did not satisfy the null hypothesis. With the aggregation of two years of test data, the sample size for each distribution tested increases.

Sixteen of the 32 cases where the null hypothesis was not satisfied were for vehicles with odometer readings greater than 85,000 miles. Two particular group/pollutant combinations gave the largest numbers of distributions failing to satisfy the null hypothesis: (1) old technology groups 8 and 9 for HC and CO, and (2) old technology group 11 for HC and NOx.

A final level of aggregation took all the measurement years for each technology group and age classification. This gave 60 total distributions, of which 39 satisfied the null hypothesis at the 0.05 level. Nine of the 21 distributions that did not satisfy the null hypothesis were for distances greater than 85,000 miles. Details of the failed distributions are shown in Table 4-14.

Table 4-13
Description of Two-Year Data Sets Not Satisfying Null Hypothesis of χ^2 Test at 0.05 Significance Level
Not Shown: 142 of 174 Data Sets Satisfying Null Hypothesis

Species	Old Technology Group	Test Years	Mileage Group	Sample Size	χ^2 Result	
					Test Statistic	Probability
HC	6	93-94	85k+	2	-	.0067
	8	89-90	85k+	69	51.40	2x10$^{=10}$
		93-94	66-85k	11	-	.0009
		93-94	85k+	23	18.55	.0009
	9	89-90	45-66k	36	22.32	.0002
		89-90	85k+	16	49.50	5x10^{-10}
		91-92	85k+	9	-	.0006
	11	89-90	45-66k	31	14.35	.0060
		89-90	85k+	43	15.90	.0032
		93-94	45-66k	14	22.24	.0002
		93-94	85k+	21	12.00	.0174
	13	91-92	22-45k	9	-	.0152
CO	6	87-88	66-85k	9	-	.0058
		93-94	85k+	2	-	.0026
	8	89-90	85k+	69	19.64	.0006
		93-94	45-66k	11	-	.0081
		93-94	66-85k	11	-	.0203
		93-94	85k+	23	12.45	.0143
	9	87-88	22-45k	17	12.11	.0166
		89-90	45-66k	36	23.12	.0001
		89-90	85k+	16	99.55	1x10^{-20}
		91-92	85k+	9	-	.0009
	11	All pass for this pollutant and technology group combination				
	13	All pass for this pollutant and technology group combination				
NOx	6	89-90	85k+	20	11.87	.0184

	8	93-94 93-94	66-85k 85k+	11 23	- 12.55	.0208 .0137
	9	87-88 89-90	0-22k 22-45k	15 11	24.02 -	$8x10^{-5}$ .0007
	11	89-90 91-92 91-92 91-92 93-94	85k+ 45-66k 66-85k 85k+ 66-85k	43 28 30 29 10	24.58 18.90 22.17 11.67 -	$6x10^{-5}$ .0008 .0002 .0200 $9x10^{-5}$
	13	All pass for this pollutant and technology group combination				

Note: The "probability" column entry represents the probability that the test statistic would be observed if the two distributions were actually the same. The test statistic column contains the result of the computation in equation [1] where the total sample was greater than 12 vehicles.

Table 4-14

Description of Six-Year Data Sets Not Satisfying Null Hypothesis of χ^2 Test at 0.05 Significance Level

Not Shown: 39 of 60 Data Sets Satisfying Null Hypothesis

Species	Old Technology Group	Test Years	Mileage Group	Sample Size	χ^2 Result	
					Test Statistic	Probability
HC	6	87-94	66-85k	27	63.05	$7x10^{-13}$
	8	87-94 87-94	66-85k 85k+	104 168	14.20 59.29	.0067 $4x10^{-12}$
	9	87-94	45-66k	104	18.90	.0008
	11	87-94	85k+	112	29.94	$5x10^{-6}$
	13	All pass for this pollutant and technology group combination				
CO	6	87-94	66-85k	27	51.51	$2x10^{-10}$
	8	87-94	85k+	168	31.73	$2x10^{-6}$

	9	87-94 87-94 87-94	22-45k 45-66k 85k+	82 104 37	16.25 10.60 121.22	.0027 .0315 $3x10^{-25}$
	11	87-94	85k+	112	23.37	.0001
	13	All pass for this pollutant and technology group combination				
NOx	6	87-94 87-94	66-85k 85k+	27 55	54.92 10.60	$3x10^{-11}$ .0314
	8	87-94	85k+	168	11.85	.0185
	9	87-94 87-94 87-94	0-22k 45-66k 85k+	17 104 37	19.42 10.47 18.44	.0007 .0332 .0010
	11	87-94 87-94 87-94	22-45k 66-85k 85k+	95 87 112	12.11 24.89 20.47	.0166 $5x10^{-5}$ .0004
	13	87-94	22-45k	23	10.45	.0033
Note: The "probability" column entry represents the probability that the test statistic would be observed if the two distributions were actually the same. The test statistic column contains the result of the computation in equation [1] where the total sample was greater than 12 vehicles.						

The aggregation of the data into larger groupings of test years was done to see if larger sample sizes would improve the matching between the EMFAC2000 predictions and the surveillance data. However, just the opposite effect occurred. As the data were aggregated into more than individual year groups, the fraction of distributions that matched the EMFAC2000 predictions declined from 85.8% for one test year to 81.6% for two test years to 65.0% for all test years. Because the EMFAC2000 predictions are for individual years and because the small sample procedure is able to provide results regardless of the sample size, the comparison should rely on the single-year results.

In order to determine the direction of the difference for the distributions that did not satisfy the null hypothesis, a simple index to measure the degree of poorly performing vehicles was constructed. The index was computed by assigning each regime a score based on the midpoint of the CALIMFAC emission boundaries for the regime. For example, the very high CO emission regime, with emissions between 6 to 10 times the FTP standard, was assigned a score of 8. The scores for all pollutants and regimes are shown in Table 4-15.

Table 4-15 Scores Used in "Dirtiness" Index for Vehicle Distributions					
Species	Normal	Moderate	High	Very High	Super
HC	0.75	1.5	3.5	7	15
CO	0.75	1.5	4	8	15
NOx	0.75	1.5	2.5	3.5	6

The index is computed from the scores in Table 4-15 by the following equation:

$$\begin{aligned}\text{Index} = [\ & s_{N,species} \text{ x (Percentage of Normals)} \\ & + s_{M,species} \text{ x (Percentage of Moderates)} \\ & + s_{H,species} \text{ x (Percentage of Highs)} \\ & + s_{V,species} \text{ x (Percentage of Very Highs)} \\ & + s_{S,species} \text{ x (Percentage of Supers)}\ ] / s_{S,species}\end{aligned}$$

This would range from a low score* for 100% normal vehicles to 100 for 100% super emitters. This index was computed for both the CALIMFAC distribution and the surveillance data in the cases where the null hypothesis was not satisfied. In most of the cases, the index was higher for the surveillance fleet than for the CALIMFAC cases. This indicates that the cases where there is a significant difference between the CALIMFAC predictions and the surveillance data are cases where CALIMFAC is underpredicting the actual emissions. The details of this comparison are given in Table 4-16. The occurrence of dirtier distributions is truer for HC and CO than it is for NOx. The one exception to this general trend was for old technology group 13. The vehicles from this group used in CALIMFAC comparisons were generally newer cars with no vehicles over 66,000 miles. The 94 vehicles in the surveillance data had no very high or super regime vehicles for HC, two very high and no super regime vehicles for CO, and two very high and five super regime vehicles for NOx. For this technology group, all the distributions that did not meet the statistical criterion for similarity had cleaner emissions than the CALIMFAC predictions.

Table 4-16 Number of Different Cases That Have "Dirtier" Vehicles in Surveillance Fleet than in EMFAC2000 Predictions Compared to Total Cases with Different Distributions						
Years of Test Data in Group	HC		CO		NOx	
	Total Cases	Dirty Cases	Total Cases	Dirty Cases	Total Cases	Dirty Cases

*The lowest scores are 100 s_N/s_S for each pollutant. These minimum values are 5 for HC and CO and 12.5 for NOx.

One year	18	16	11	11	13	8
Two years	12	11	10	9	10	5
Six years	5	5	6	6	10	6
In this table, "Total Cases" refers to total cases with different distributions. For example, the one-year data set for HC had 18 cases where the surveillance data had a significantly different distribution than the EMFAC2000 predictions (out of a possible 101 cases tested). Of these 18 cases, 16 had a "dirtier" distribution in the surveillance data than in the CALIMFAC predictions.						

The comparisons made here included data from the "Surveillance-9" study (2S87C1). These data were also used in the derivation of CALIMFAC regime boundaries and regime growth functions. It was included in the comparisons here as a check on how good a match could be expected by the χ^2 comparison. The Surveillance-9 data were all taken in 1987 and 1988; no other data were taken in those years. The single-year comparisons in Table 3 show that four technology group/mileage/pollutant combinations do not show a match between CALIMFAC predictions and the Surveillance-9 data. This shows that the statistical fits to the data used in CALIMFAC have some residual error and it is not surprising that data, which were not used in the fit, do not match the CALIMFAC data completely. However, the fits for the Surveillance-9 data are better than those for other years. If the 43 distributions that did not match were evenly distributed among the study years, the number of mismatched distributions for 1987 and 1988 would be expected to be about ten instead of the four actually found.

Based on the comparison, the following conclusions were reached:

- CALIMFAC predictions are consistent with surveillance data for 86% of the cases evaluated.
- Significant differences exist and particular attention was paid to the following technology groups in subsequent development of emission rates for EMFAC2000:

    - vehicles with more than 85,000 miles;
    - HC and CO emissions from old technology group 8;
    - all emissions from old technology group 9; and
    - HC and NOx emissions from old technology group 11.

- In general, the differences occurred because the vehicles in the sample were dirtier than the CALIMFAC predictions. The exception to this was old technology group 13.

Having determined that there were several shortcomings of the then CALIMFAC model, staff decided on data to be used in updating the EMFAC2000 model. The model is driven by data from vehicles that have not been subject to I&M programs. Further, even the data from CARB's surveillance projects could not be simply added to the "EMFAC2000 master" data set without making sure that the data was indeed representative of the failure rates found during BAR's random roadside surveys. The following section details the data used in EMFAC2000.